Leg Posture Correction System for Physical Education Students Based on Multimodal Information Processing

Lei Yang[1(✉)] and Yueguo Jia[2]

[1] Xi'an Medical College, Xi'an 710021, China
yl19810920@126.com
[2] Tianjin Public Security Professional College, Tianjin 300382, China

Abstract. The traditional leg posture correction system has the problem that the joint points are arranged in reverse order and connected incorrectly, which affects the accuracy of posture recognition. In response to this problem, this research designed a leg posture correction system for students in physical education class based on multi-modal information processing. In the hardware part of the system, a signal conditioning circuit is used to filter, amplify, and sample the input signal, and an operational amplifier with a zero-adjusting terminal is used in conjunction with a D/A converter to realize the zero-adjustment of the circuit. In the software part of the system, after segmenting the depth image of the scene object containing the viewpoint, establish a database of the leg pose of the students in physical education class, then fuse the data vector and use the multi-modal information processing model to recognize the leg pose, and use the recognition result as the misrecognition probability matrix model The input to realize the intelligent error correction of the wrong leg posture. The experimental results show that under the test condition of 100 users, the positioning accuracy of the leg parts of the system in this paper is as high as 86.85%, which proves that it has a good application effect.

Keywords: Multimodal information processing · PEclass · Leg posture · Attitude correction

1 Introduction

Speaking of sports, from the educational level, it is a teaching process to promote sports, improve human quality, preach the skills and methods of physical exercise, and promote the all-round development of morality and will quality. At the same time, physical education is the experience of cultivating and shaping the body, and it is an important means to cultivate people into wholehearted development. The goal and key of physical education is to pursue standardized sports indicators and scientifically guide training and teaching.

The introduction of human posture recognition into sports can provide a new idea and scheme for sports training and action evaluation. Attitude correction system is of

W. Fu and G. Sun (Eds.): eLEOT 2022, LNICST 453, pp. 91–102, 2022.
https://doi.org/10.1007/978-3-031-21161-4_8

great significance to improve motion quality. Human pose recognition is defined as the localization of human joints in images or videos. It is also defined as searching for specific poses in the space composed of all joint poses.

At present, the research on motion attitude error correction system has achieved some research results. In reference [1], a sprint leg posture monitoring system based on inertial sensor is proposed. The system uses wireless sensor network to design distributed sprint leg posture information acquisition nodes, installs pressure sensors and inertial sensors in athletes' legs, and carries out information fusion and association mining according to the output oscillation amplitude of the sensors, so as to realize the intelligent monitoring of sprint leg posture data. A gait control system based on quantitative fusion of inertial attitude parameters is proposed in reference [2]. The system constructs the kinematic model of gait, models the constraint parameters of gait control, collects the attitude parameters with position and attitude sensors such as gyroscope and accelerometer, fuses the inertial attitude parameters with extended Kalman filter and inputs them into the real-time control actuator to realize stable gait control. A neural network based attitude estimation error compensation system is proposed in reference [3]. With the help of the nonlinear mapping ability of BP neural network, the attitude error compensation model between sensor output and attitude estimation error is established, which effectively improves the detection accuracy. However, the above-mentioned posture error correction systems can all realize the function of assisting students in physical training, but there is also the problem of incorrect connection of joint points in reverse order.

Multi-modal information processing technology can effectively fuse the information obtained by multiple and multiple types of sensors to form a high-level information expression form. The multi-modal information fusion technology comprehensively uses signal processing, mathematical statistics and artificial intelligence and other related theories to fuse the multi-modal information provided by sensors distributed in different positions, different types and different states of the attitude correction system, and at the same time The redundancy and contradiction that may exist between multi-modal information are eliminated, and the information obtained by each sensor is fully utilized, while achieving their respective advantages and complementing each other, and finally forming an overall, complete, and consistent of their own state and the surrounding environment The perceptual description of the system provides a basis for the self-positioning, intelligent decision-making and other related needs of gesture recognition.

Based on the above analysis, this article designs a new leg posture correction system for students in physical education class based on multi-modal information processing to help students understand the difference between their current movements and standard movements in time, and make timely repairs based on the training suggestions fed back by the system., To achieve high-quality sports effects. In this study, the signal conditioning circuit is used to filter, amplify and sample the input signal, which fundamentally improves the accuracy of subsequent error correction. Secondly, the multi-modal information processing model is used to identify the leg posture, and the recognition results are used as the input of the error probability matrix model to realize the intelligent error correction of the leg posture, so as to improve the positioning accuracy of the leg posture.

2 System Hardware Design

The leg posture correction system of physical education students must measure different postures, so the signal conditioning circuit is used to filter, amplify and sample the input signal. Because each measurement node contains four measurement channels, the single chip microcomputer can only collect data by polling. In order to meet the requirements of acquisition synchronization and accuracy, each input strain measurement signal is designed with an independent bridge circuit, and each signal is successively sent to the A/D converter port by a high-speed analog switch.

The driver chip selected in this article is Sn74alvc164245 from TI. The chip can amplify the drive capacity of input and output signals, increase its drive capacity, and can drive the subsequent circuit to make it work normally. When selecting the device, pay attention to the channel switching time and signal transmission delay, and try to shorten the channel switching time [4].

During the load calibration test, the system should be zeroed before measuring the strain to ensure the accuracy of the measurement. The realization of the zero-adjusting circuit can be realized by using an operational amplifier with a zero-adjusting terminal and a D/A converter. The strain bridge is realized by an unbalanced electric bridge, and considering the problem of low power consumption, it is realized by a constant voltage bridge. In the communication serial module, the upper computer is the PC, and the lower computer is the smallest DSP system. The PC transmission data level is 1.5 V, the DSP data transmission I/O port level is TTL level, and the TTL level is “1” The characteristic voltages of and “0” are 2.4 V and 0.4 V respectively, which are suitable for data transmission within the board. There is no direct transmission between PC and DSP. In order to solve the level inequality, a serial port conversion chip is needed to convert the level. Choose MAX3232 chip here to switch between CMOS level and TTL level. The 9th and 10th pins link the receiving data bit and the sending data bit of the DSP, so the DSP transmits the data directly to the upper computer. Due to the limitation of the printed circuit board area, the strain bridge adopts a 1/4-type semi-equal-arm Wheatstone bridge. Secondly, choose a three-wire external sensor to weaken the bridge arm imbalance caused by the parasitic resistance of the lead cable and reduce the measurement error in the measurement process.

The schematic diagram of the strain bridge circuit is shown in Fig. 1.

The communication protocol adopts the RS232 protocol, which is an asynchronous communication protocol, and adopts an unbalanced transmission method, which is the so-called single-ended communication. The check method is changed to cyclic redundancy check to reduce the data packet loss rate and protect the data from loss.

In the process of system communication, the enhanced feature FIFO function is used for data transmission, which basically does not occupy the cycle waiting time of the system, and write data directly to the BUFF and send it directly. The master-slave clock configuration of the signal acquisition and transmission board can be configured through the serial communication interface or the GPIO interface. When the acquisition and transmission board receives a command to configure it as the main board, the master signal acquisition and transmission board only outputs clock signals at this time. For the slave board, on the one hand, it receives the clock signal transmitted from the upper level, and on the other hand, transmits the received clock signal to the next level.

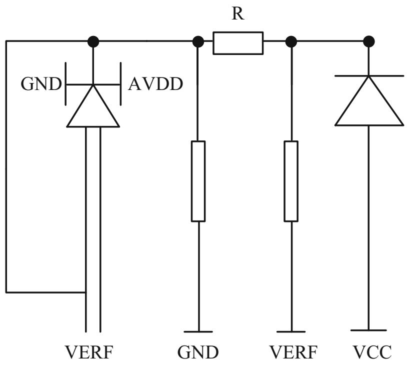

Fig. 1. Structure diagram of strain bridge circuit

In the FIFO mode, the transfer of information is based on multiple frames forming a transfer unit to transfer data, rather than transmitting the data frame by frame. FIFO means first-in-first-out, that is, the first received data is read first. The 16-level depth can only save the first 16-level data. If the depth is exceeded, data overflow will occur, and the data will be lost. Use software Can deal with overflow.

3 System Software Design

3.1 Establishing a Database of Leg Postures of Students in Physical Education Class

The design of physical education students' leg posture database is an interactive way based on machine vision. In this paper, the commonly used static and dynamic leg postures in line with people's cognitive laws and behavior habits are counted. The database includes depth map, color map and position data of leg trajectory.

The image acquisition process will be affected by light intensity, external conditions and equipment functions. In the image transmission process, it will be affected by noise (common thermal noise), resulting in the deformation of the collected image effect, which will affect the accuracy of the leg posture error correction system.

This article will explain the process of database collection and establishment based on Kinect. The low-cost color depth (RGB-D) camera Kinect released by Microsoft has created a new revolution in gesture recognition by providing high-quality depth images to solve complex background and lighting changes. Kinect sensor integrates a variety of advanced sensing hardware. The most worth mentioning is that it contains a depth sensor, a color camera and a microphone array, which can provide full-body 3D motion capture, facial recognition and voice recognition. Kinect can directly track the position information of the leg joint points, so this article selects the leg joint point as the tracking center based on Kinect.

When processing the image, we should make corresponding preprocessing to enhance the image quality, and eliminate the ineffective information in the image to ensure the accuracy and precision of the image. On the whole, median filtering is a non-linear method, which is used for image recognition and smoothing filtering processing skills. Then, all pixels in the area near the image are arranged according to the order of gray values, and the middle value is found as the pixel value to be output [5].

Image segmentation is a key step in establishing a database. The segmented image is the object of leg gesture recognition and the source of the entire interactive process. Compared with color images, depth images are based on adding depth maps. The human pose tracking problem can essentially be regarded as an extension of the human pose estimation problem in static images in the time domain, so the above-mentioned method can also be directly used to solve the pose estimation in video sequence images. In 3D computer graphics, a depth map is an image or image channel that contains information about the distance to the surface of the scene object of the viewpoint. Therefore, the database of this article uses depth images. Kinect can obtain the depth map of the user's whole body, so it is only necessary to separate the leg area from the body as a whole.

The approach taken in this article is to select the area at the same depth as the leg as the destination and save the destination directly in this article file. Because the gray value of the boundary of the region will lead to discontinuous phenomenon, the method of detecting these regions can find the edge of the region to be detected, and at the same time, it also includes the segmentation of the target image range, so as to get the depth image of the leg posture [6].

3.2 Recognizing Leg Posture Based on Multi-modal Information Processing

The basic process of multimodal information fusion is to make full and efficient use of the information obtained by multiple sensors, obtain redundant or complementary information in space or time, organically combine through the fusion algorithm, form a consistent interpretation or description of the measured object or perceived object, and then obtain better performance than a single sensor system.

The multimodal model not only contains the spatial information of a single frame image, but also considers the implicit time information between adjacent images, so that it can effectively extract the spatio-temporal features in the video sequence, so as to obtain a recognition effect much higher than that of convolutional neural network [7]. In the feature layer fusion, the observation information of each sensor completes the feature extraction, obtains the feature vectors from each sensor, and then fuses these feature vectors to generate joint feature vectors.

Each feature dimension in the fusion feature vector is not unified. Firstly, the feature vectors of each group are associated to ensure that the feature vectors participating in the fusion come from the same target. Considering that the leg is the main moving object in the pose sequence, it often has high optical flow amplitude. Therefore, by setting an appropriate threshold, the corresponding leg image blocks can be filtered out from the image. Recognition according to the input of different modes, including sensing and recognition of sensor information, attitude information and voice information. The fused feature can be a higher dimensional feature vector connected by each group of feature vectors, or a new type of feature vector combined by each group of feature vectors.

Due to the uncertainty of the proportion of the leg region, the number of elements in the set is not fixed. The intensity of leg motion in the current image is expressed by calculating the average optical flow amplitude of the leg image block [8]. The average optical flow amplitude is calculated as follows:

$$\overline{a} = \frac{b_i}{n} \tag{1}$$

In formula (1), $\overline{a}$ represents the average optical flow amplitude; b_i represents the element in the optical flow amplitude set; i represents the sequence number of the element; n represents the total number of elements in the set. In the multi-modal information fusion interaction strategy, different nodes represent different states, and each state represents the intention of a user. The multi-modal model will sample each image sequence, and then use the obtained key frame sequence as input for subsequent processing. When segmenting the sequence, the length of the sub-sequence should be reasonably controlled to facilitate the sampling operation. The default key frame sequence length is 16 frames [9], and its advantages are good real-time performance, strong fault tolerance, and small dependence on original data. After the sequence segmentation process, each video is divided into four sequences, the sequences are independent of each other, and they are sequentially sent to the classification model for classification training.

SVM is a class II linear classifier that searches for the maximum interval in the feature space, and its objective function can be expressed as:

$$\begin{cases} f = \min \frac{1}{2}\|\beta\|^2 \\ Q_j\left(\beta^T P_j + \delta\right) \geq 1 \end{cases} \tag{2}$$

In formula (2), f represents the objective function; β represents the geometric distance of the classification interval; Q_j and P_j represent the output and input vectors of the sample data; T represents the matrix transposition; j represents the serial number of the sample data; δ represents the hyperplane parameter. According to different operations, the system enters different states, and feeds back the expression effect of state semantics, and then conducts voice navigation and broadcast according to the system's intentions.

The characteristics of various data type sequences are fully learned from the multi-modal data processing model. Each data type corresponds to a trained multimodal neural network, and then the output data of a layer of the network are fused in the test stage. Finally, the human posture and action recognition is realized based on the fused data.

3.3 Design Error Correction Model of Leg Posture

In order to solve the problem of leg posture confusion caused by the wrong arrangement and connection of joint points in reverse order, a leg posture error correction model based on probability statistics is proposed according to the theorem of large numbers.

The leg posture error correction model has two preconditions: the first premise is that different types of joint points have different confidence, and the second premise is that the candidate joint points predicted by the network need to be located in the human body area. Confusion matrix is usually called matching matrix in unsupervised learning. The columns of the matrix represent examples of prediction classes and the rows represent examples of actual classes, so the accuracy of the algorithm can be measured by confusing some metrics in the matrix [10].

The repeated counting of joint points means that the joint points return to the corresponding and symmetrical joint area, resulting in two kinds of symmetrical joint points in the same joint range. The inverse arrangement of joint points means that the position of lower joint points is higher than that of upper joint points. Each row represents the predicted value and each row represents the actual category. The total number of each column represents the predicted quantity of the category, and the total number of each row represents the real quantity of the category. The values in each column represent the number of actual data predicted for this class. The misjudgment probability matrix can be expressed as:

$$W_{\alpha\gamma} = \frac{z_{\alpha\gamma}}{\sum_{\gamma=0}^{3} z_{\alpha\gamma}} \tag{3}$$

In formula (3), $W_{\alpha\beta}$ represents the probability matrix of misjudgment; α represents the recognition result; γ represents the source label; $z_{\alpha\gamma}$ represents the recognition result is α, which actually comes from the number of labels γ.

Misjudgment probability model is to optimize the recognition results of leg posture model based on convolutional neural network. It should not only ensure the correctness of the original recognition results, but also correct the wrong leg posture of students, and control the new category with random numbers. When the number of similar joint points is greater than two, the solution is similar to that of only two similar joint points.

For example, when there are three repeated counting problems of knee joint points, first divide the knee joint points into two groups, and there are three groups with two knee joints; Secondly, the hip joint was connected with three groups of knee joint points; Then it is processed according to the principle that there are only two joint points of the same kind. When the hip joint and knee joint are accurately positioned, the knee joint and ankle joint are treated in the same way from the correct knee joint point. The structure of leg posture error correction model is shown in Fig. 2.

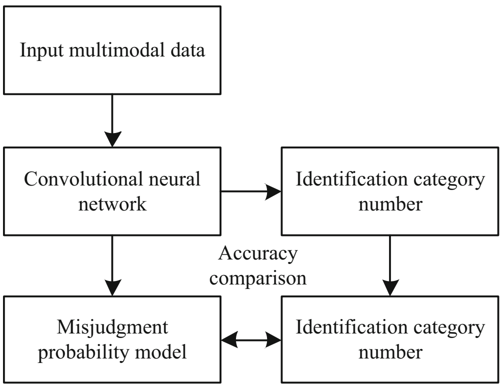

Fig. 2. Leg posture error correction model

The multi-modal image data is recognized by the convolutional neural network model, and the recognition result is used as the input of the misrecognition probability matrix model. The error correction model is also composed of a convolutional layer and a pooling layer, and each neuron is only responsible for processing the local information of the feature map of the previous layer. Assuming that the recognition result is 01, 01 is the input parameter of the error correction algorithm. After calling the model, a new category number will be generated. The number of such recognition results is counted to calculate the recognition rate. Based on the confusion matrix of the actual class and the prediction class, the model establishes the misjudgment probability function of the prediction class and the actual class, so as to realize the intelligent error correction of the wrong leg posture.

So far, the design of the leg posture error correction system for students in physical education class based on multi-modal information processing has been completed. The specific system structure is shown in Fig. 3.

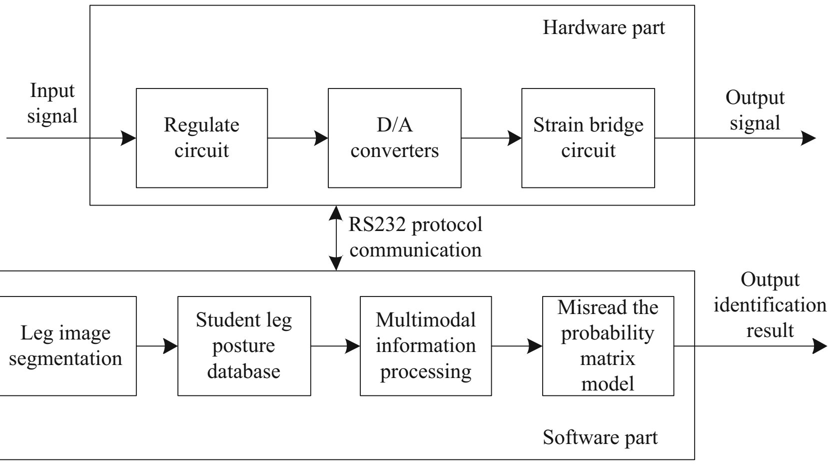

Fig. 3. Overall system structure drawing

4 System Test and Result Analysis

In order to verify the practical application performance of PE Students' leg posture error correction system based on multimodal information processing, the following experiments are designed.

4.1 Experiment Preparation

The hardware environment of this system test is as follows:

CPU: i7-8550U;

GPU: MX150, memory 8G, video memory 4G;
Experimental external equipment: Kinect 2.0, smart equipment;
Input data: student's leg posture sensor signal.

The system software platform is mainly built in the Visual Studio 2015 environment. In the development process, the open source computer vision library OpenCV for deep neural networks, Microsoft's MFC interface library, CUDA architecture and GPU acceleration library cuDNN were used. The system uses C++ for development and processes the files, data types, text streams and other objects involved in the program.

4.2 System Performance Test Results and Analysis

In order to verify the application effect of the proposed system, it is compared with the attitude correction system based on inertial attitude parameter quantization fusion (traditional system 1) and the attitude correction system based on neural network compensation algorithm (traditional system 2).

The evaluation index selected in this experiment is the percentage of correctly positioned parts. For each joint point, when the distance between the predicted position and the true position is less than a given threshold, the joint point is correctly positioned. When the two joint points corresponding to the two ends of the human body part are correctly positioned, the part is correctly positioned. Under the condition that the number of test users is 10, 100 and 500, the positioning accuracy of the leg parts of each posture error correction system is tested respectively. The results are shown in Tables 1, 2 and 3.

Table 1. Test results with 10 users

Number of experiments	Accuracy of leg posture positioning/%		
	System of this paper	Traditional system 1	Traditional system 2
1	91.44	85.74	81.04
2	92.88	84.87	82.45
3	90.66	86.95	84.58
4	89.55	85.66	85.66
5	88.27	84.52	83.22
6	91.04	87.83	82.05
7	90.10	85.41	83.13
8	92.55	86.85	81.52
9	93.22	84.53	80.81
10	92.94	84.02	82.90

Under the test condition of 10 users, the maximum positioning accuracy of leg parts in this system is 93.22%, which is 5.39% and 7.56% higher than that of Traditional System 1 and Traditional System 2, respectively. Therefore, when the number of users is 10, compared with the two traditional systems, system of this paper can locate the leg posture more accurately.

Table 2. Test results with 100 users

Number of experiments	Accuracy of leg posture positioning/%		
	System of this paper	Traditional system 1	Traditional system 2
1	88.41	81.49	78.04
2	87.84	82.56	79.48
3	85.55	83.67	80.58
4	86.68	81.88	79.67
5	85.26	80.26	78.24
6	84.51	81.35	77.58
7	85.82	80.54	78.85
8	87.66	80.01	79.66
9	88.53	82.48	80.22
10	88.20	82.82	78.34

Under the test condition of 100 users, the maximum positioning accuracy of leg parts in this system is 88.53%, which is 4.86% and 7.95% higher than that of Traditional System 1 and Traditional System 2, respectively. Therefore, when the number of users is 100, compared with the two traditional systems, system of this paper can effectively improve the positioning accuracy of leg posture.

When the number of users is 500, the accuracy of leg posture positioning of the system is up to 85.56%, that of traditional system 1 is 78.15%, and that of traditional system 2 is 74.64%. In contrast, the system in this paper can obtain the leg posture localization results more accurately.

Based on the above results, it can be seen that the leg posture correction system designed in this paper has obvious advantages in the positioning accuracy of the key points, and can better achieve high-precision human posture estimation and error correction tasks.

Table 3. Test results with 500 users

Number of experiments	Accuracy of leg posture positioning/%		
	System of this paper	Traditional system 1	Traditional system 2
1	85.44	77.44	72.62
2	84.71	76.27	74.64
3	83.50	75.58	73.81
4	82.63	74.65	71.56
5	84.25	75.36	72.32
6	85.56	76.02	72.25
7	83.82	78.15	73.52
8	83.95	77.24	72.10
9	84.67	76.52	71.36
10	83.21	75.81	72.08

5 Conclusion

Human pose estimation in natural images has always been a hot topic in the field of computer vision. It has a good application prospect in many fields, such as behavior recognition, human-computer interaction, motion analysis and so on. This paper designs a leg posture error correction system for physical education students from the perspective of multimodal information processing. The system can accurately locate the leg joints and parts, and improve the accuracy of pose recognition. The research of this paper mainly aims at the problem of single person pose estimation in natural images, but in fact, it is more common to have multiple human objects in natural images at the same time. Therefore, further research on multi person human pose in natural images is more in line with the needs of the market.

References

1. Yang, L.: Monitoring method of leg posture of sprinters based on inertial sensor. J. Hebei North Univ. (Nat. Sci. Ed.) **36**(5), 42–47 (2020)
2. Lu, X., Huang, W., Huang, P.: Bionic gait control of robot leg based on quantization fusion of inertial attitude parameters. Transducer Microsyst. Technol. **38**(3), 43–46 (2019)
3. Sun, Y., He, S., Xu, X., et al.: Application of neural network compensation algorithm in MEMS-based attitude measurement. Appl. Res. Comput. **36**(9), 2696–2699 (2019)
4. Wang, L., Zhang, Z., Niu, Q., et al.: Design of human body posture detection system based on MPU9250 and MS5611. Chin. J. Electron Dev. **42**(4), 978–983 (2019)
5. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. Neurocomputing **458**, 615–625 (2021)
6. Liu, Y., Hui, H., Lu, Y., et al.: Remote human body posture monitoring system based on smart phone terminal. J. Chin. Inert. Technol. **27**(6), 713–718 (2019)

7. Liu, S., Wang, S., Liu, X., et al.: Fuzzy detection aided real-time and robust visual tracking under complex environments. IEEE Trans. Fuzzy Syst. **29**(1), 90–102 (2021)
8. Zhu, H., Yin, J., Feng, W., et al.: Research and application of a lightweight real-time human posture detection model. J. Syst. Simul. **32**(11), 2155–2165 (2020)
9. Liu, S., et al.: Human memory update strategy: a multi-layer template update mechanism for remote visual monitoring. IEEE Trans. Multimedia **23**, 2188–2198 (2021)
10. Lu, J.: Attitude recognition simulation of data glove based on feature point set matching strategy. Comput. Simul. **38**(4), 403–407 (2021)